Approximation Algorithms for the Joint Replenishment Problem with Deadlines*

Marcin Bienkowski[1], Jaroslaw Byrka[1], Marek Chrobak[2], Neil Dobbs[3], Tomasz Nowicki[3], Maxim Sviridenko[4], Grzegorz Świrszcz[4], and Neal E. Young[2]

[1] Institute of Computer Science, University of Wrocław, Poland
[2] Department of Computer Science, University of California at Riverside, USA
[3] IBM T.J. Watson Research Center, Yorktown Heights, USA
[4] Department of Computer Science, University of Warwick, UK

Abstract. The Joint Replenishment Problem (JRP) is a fundamental optimization problem in supply-chain management, concerned with optimizing the flow of goods over time from a supplier to retailers. Over time, in response to demands at the retailers, the supplier sends shipments, via a warehouse, to the retailers. The objective is to schedule shipments to minimize the sum of shipping costs and retailers' waiting costs.

We study the approximability of JRP with deadlines, where instead of waiting costs the retailers impose strict deadlines. We study the integrality gap of the standard linear-program (LP) relaxation, giving a lower bound of 1.207, and an upper bound and approximation ratio of 1.574. The best previous upper bound and approximation ratio was 1.667; no lower bound was previously published. For the special case when all demand periods are of equal length we give an upper bound of 1.5, a lower bound of 1.2, and show APX-hardness.

Keywords: Joint replenishment problem with deadlines, inventory theory, linear programming, integrality gap, randomized rounding, approximation algorithm.

1 Introduction

The Joint Replenishment Problem with Deadlines (JRP-D) is an optimization problem in inventory theory concerned with optimizing a schedule of shipments of a commodity from a supplier, via a warehouse, to satisfy demands at m retailers (cf. Figure 1). An instance is specified by a tuple $(C, c, \mathcal{D})$ where $C \in \mathbb{Q}$ is the *warehouse ordering cost*, each retailer $\rho \in \{1, 2, \ldots, m\}$ has *retailer ordering cost* $c_\rho \in \mathbb{Q}$, and $\mathcal{D}$ is a set of n *demands*, where each demand is a triple (ρ, r, d), where ρ is a retailer, $r \in \mathbb{N}$ is the demand's release time and $d \in \mathbb{N}$ is its deadline. The interval $[r, d]$ is the demand *period*. Without loss of generality, we assume $r, d \in [2n]$, where $[i]$ denotes $\{1, 2, \ldots, i\}$.

* Full version available at [4]. Research supported by MNiSW grant number N N206 368839 (2010-2013); NSF grants CCF-1217314, CCF-1117954, OISE-1157129; EPSRC grants EP/J021814/1 and EP/D063191/1; FP7 Marie Curie Career Integration Grant; and Royal Society Wolfson Research Merit Award.

F.V. Fomin et al. (Eds.): ICALP 2013, Part I, LNCS 7965, pp. 135–147, 2013.

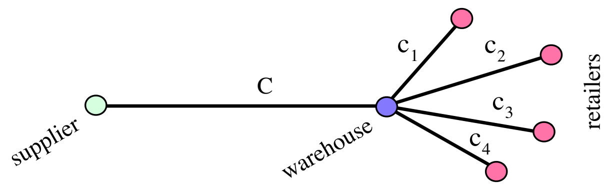

Fig. 1. An instance with four retailers, with ordering costs as distances. The cost of an order is the weight of the subtree connecting the supplier and the involved retailers.

A solution (called a *schedule*) is a set of *orders*, each specified by a pair (t, R), where t is the time of the order and R is a subset of the retailers. An order (t, R) *satisfies* those demands (ρ, r, d) whose retailer is in R and whose demand period contains t (that is, $\rho \in R$ and $t \in [r, d]$). A schedule is *feasible* if all demands are satisfied by some order in the schedule.

The *cost* of order (t, R) is the ordering cost of the warehouse plus the ordering costs of respective retailers, i.e., $C + \sum_{\rho \in R} c_\rho$. It is convenient to think of this order as consisting of a warehouse order of cost C, which is then joined by each retailer $\rho \in R$ at cost c_ρ. The cost of the schedule is the sum of the costs of its orders. The objective is to find a feasible schedule of minimum cost.

Previous Results. The decision variant of JRP-D was shown to be strongly $\mathbb{NP}$-complete by Becchetti et al. [3]. (They considered an equivalent problem of packet aggregation with deadlines on two-level trees.) Nonner and Souza [12] then showed that JRP-D is $\mathbb{APX}$-hard, even if each retailer issues only three demands. Using the primal-dual method, Levi, Roundy and Shmoys [9] gave a 2-approximation algorithm. Using randomized rounding, Levi et al. [10,11] (building on [8]) improved the approximation ratio to 1.8; Nonner and Souza [12] reduced it further to 5/3. These results use a natural linear-program (LP) relaxation, which we use too.

The randomized-rounding approach from [12] uses a natural rounding scheme whose analysis can be reduced to a probabilistic game. For any probability distribution p on $[0, 1]$, the integrality gap of the LP relaxation is at most $1/\mathcal{Z}(p)$, where $\mathcal{Z}(p)$ is a particular statistic of p (see Lemma 1). The challenge in this approach is to find a distribution where $1/\mathcal{Z}(p)$ is small. Nonner and Souza show that there is a distribution p with $1/\mathcal{Z}(p) \leq 5/3 \approx 1.67$. As long as the distribution can be sampled from efficiently, the approach yields a polynomial-time $(1/\mathcal{Z}(p))$-approximation algorithm.

Our Contributions. We show that there is a distribution p with $1/\mathcal{Z}(p) \leq 1.574$. We present this result in two steps: we show the bound $e/(e-1) \approx 1.58$ with a simple and elegant analysis, then improve it to 1.574 by refining the underlying distribution. We conjecture that this distribution minimizes $1/\mathcal{Z}(p)$. This shows that the integrality gap is at most 1.574 and gives a 1.574-approximation algorithm. We also prove that the LP integrality gap is *at least* 1.207 and we provide a computer-assisted proof that this gap is at least 1.245. (As far as we know, no explicit lower bounds have been previously published.)

For the special case when all demand periods have the same length (as occurs in applications where time-to-delivery is globally standardized) we give an upper bound of 1.5, a lower bound of 1.2, and show $\mathbb{APX}$-hardness.

Related Work. JRP-D is a special case of the Joint Replenishment Problem (JRP). In JRP, instead of having a deadline, each demand is associated with a delay-cost function that specifies the cost for the delay between the times the demand was released and satisfied by an order. JRP is $\mathbb{NP}$-complete, even if the delay cost is linear [2,12]. JRP is in turn a special case of the One-Warehouse Multi-Retailer (OWMR) problem, where the commodities may be stored at the warehouse for a given cost per time unit. The 1.8-approximation by Levi et al. [11] holds also for OWMR. JRP was also studied in the online scenario: a 3-competitive algorithm was given by Buchbinder et al. [6] (see also [5]).

Another generalization of JRP involves a tree-like structure with the supplier in the root and retailers at the leaves, modeling packet aggregation in convergecasting trees. A 2-approximation is known for the variant with deadlines [3]; the case of linear delay costs has also been studied [7].

The LP Relaxation. Here is the standard LP relaxation of the problem. Let $U = \max\{d \mid (\rho, r, d) \in \mathcal{D}\}$ be the maximum deadline, and assume that each release time and deadline is in universe $\mathcal{U} = [U]$.

$$\begin{array}{lrcll} \text{minimize} & cost(\boldsymbol{x}) & = & \sum_{t=1}^{U} \left(C\, x_t + \sum_{\rho=1}^{m} c_\rho\, x_t^\rho \right) & \\ \text{subject to} & x_t & \geq & x_t^\rho \quad \text{for all } t \in \mathcal{U}, \rho \in \{1, \ldots, m\} & (1) \\ & \sum_{t=r}^{d} x_t^\rho & \geq 1 & \quad \text{for all } (\rho, r, d) \in \mathcal{D} & (2) \\ & x_t, x_t^\rho & \geq 0 & \quad \text{for all } t \in \mathcal{U}, \rho \in \{1, \ldots, m\}. & \end{array}$$

We use $\boldsymbol{x}$ to denote an optimal fractional solution to this LP relaxation.

2 Upper Bound of 1.574

The statistic $\mathcal{Z}(p)$. The approximation ratio of algorithm Round_p (defined below) and the integrality gap of the LP are at most $1/\mathcal{Z}(p)$, where $\mathcal{Z}(p)$ is defined by the following so-called *tally game* (following [12]). To begin the game, fix any *threshold* $z \geq 0$, then draw a sequence of independent samples $s_1, s_2, \ldots, s_h$ from p, stopping when their sum exceeds z. Call $z - (s_1 + s_2 + \ldots + s_{h-1})$ the *waste*. Note that, since the waste is less than s_h, it is in $[0, 1)$. Let $\mathcal{W}(p, z)$ denote the expectation of the waste. Abusing notation, let $\mathbf{E}[p]$ denote the expected value of a single sample drawn from p. Then $\mathcal{Z}(p)$ is defined to be the minimum of $\mathbf{E}[p]$ and $1 - \sup_{z \geq 0} \mathcal{W}(p, z)$.

Note that the distribution p that chooses $1/2$ with probability 1 has $\mathcal{Z}(p) = 1/2$. The challenge is to increase $\mathbf{E}[p]$ and reduce the maximum expected waste.

A Generic Randomized-Rounding Algorithm. Next we define the randomized-rounding algorithm Round_p. The algorithm is parameterized by any probability distribution p on $[0, 1]$.

For the rest of this section, fix any instance $(C, c, \mathcal{D})$ and fractional solution $\boldsymbol{x}$ of the LP relaxation. Define $\hat{U} = \sum_{t=1}^{U} x_t$. For each retailer ρ, let ω_ρ denote the piecewise-linear continuous bijection from $[0, \hat{U}]$ to $[0, \sum_{t=1}^{U} x_t^\rho]$ such that $\omega_\rho(0) = 0$, and, for each $T \in [U]$,

$$\omega_\rho(\textstyle\sum_{t=1}^{T} x_t) \;=\; \textstyle\sum_{t=1}^{T} x_t^\rho.$$

The intuition is that $\omega_\rho(z)$ is the cumulative fractional number of orders joined by retailer ρ by the time the fractional number of warehouse orders reaches z. The equations above determine ω_ρ at its breakpoints; since ω_ρ is piecewise linear and continuous, this determines the entire function. The LP inequalities (1) imply that $0 \leq \omega_\rho(z') - \omega_\rho(z) \leq z' - z$ when $z' \geq z$. That is, ω_ρ has derivative in $[0, 1]$. Here is the rounding algorithm. Recall that $\hat{U}$ denotes $\sum_{t=1}^{U} x_t$.

Algorithm $\mathsf{Round}_p(C, c_\rho, \mathcal{D}, \boldsymbol{x})$

1: Draw independent random samples $s_1, s_2, \ldots$ from p. Let $g_i = \sum_{h \leq i} s_h$.
Set *global cutoff sequence* $\boldsymbol{g} = (g_1, g_2, \ldots, g_I)$, where $I = \min\{i \mid g_i \geq \hat{U} - 1\}$.

2: For each retailer ρ independently, choose ρ's *local cutoff sequence* $\boldsymbol{\ell}^\rho \subseteq \boldsymbol{g}$ greedily to touch all intervals $[a, b]$ with $\omega_\rho(b) - \omega_\rho(a) \geq 1$.
That is, $\boldsymbol{\ell}^\rho = (\ell_1^\rho, \ell_2^\rho, \ldots, \ell_{J^\rho}^\rho)$ where ℓ_j^ρ is $\max\{g \in \boldsymbol{g} \mid \omega_\rho(g) - \omega_\rho(\ell_{j-1}^\rho) \leq 1\}$ (interpret ℓ_0^ρ as 0), and J^ρ is $\min\{j \mid \omega_\rho(\hat{U}) - \omega_\rho(\ell_j^\rho) \leq 1\}$.

3: For each $g_i \in \boldsymbol{g}$, define time $t_i \in [U]$ to be minimum such that $\sum_{t=1}^{t_i} x_t \geq g_i$.
Return the schedule $\big\{\big(t_i, \{\rho \mid g_i \in \boldsymbol{\ell}^\rho\}\big) \mid g_i \in \boldsymbol{g}\big\}$.

The idea of algorithm Round_p and its analysis are from [12]. The presentation below highlights some important technical subtleties in the proof.

Lemma 1. *For any distribution p and fractional solution $\boldsymbol{x}$, the above algorithm,* $\mathsf{Round}_p(C, c, \mathcal{D}, \boldsymbol{x})$*, returns a schedule of expected cost at most cost*$(\boldsymbol{x})/\mathcal{Z}(p)$.

Proof. Feasibility. Suppose for contradiction that the schedule does not satisfy some demand (ρ, r, d) in $\mathcal{D}$. Then (ignoring boundary cases) there are consecutive local cutoffs ℓ_j^ρ and ℓ_{j+1}^ρ equal to global cutoffs g_i and $g_{i'}$ whose times t_i and $t_{i'}$ (per Step 3) satisfy $t_i < r \leq d < t_{i'}$, and, hence, $t_i + 1 \leq r \leq d \leq t_{i'} - 1$. But, then, by Step 2 of the algorithm,

$$1 \;\geq\; \omega_\rho(\ell_{j+1}^\rho) - \omega_\rho(\ell_j^\rho) \;=\; \omega_\rho(g_{i'}) - \omega_\rho(g_i) \;>\; \sum_{t=t_i+1}^{t_{i'}-1} x_t^\rho \;\geq\; \sum_{t=r}^{d} x_t^\rho \;\geq\; 1,$$

where the last step follows from LP constraint (2), and the proper inequality in the second-to-last step follows from the minimality of $t_{i'}$ in Step 3 of the algorithm. This gives $1 > 1$, a contradiction. (The boundary cases, and the proof that Step 2 is well-defined, are similar.)

To finish the proof, for each term in the cost $C|\boldsymbol{g}| + \sum_\rho c_\rho |\boldsymbol{\ell}^\rho|$, we bound the term's expectation by $1/\mathcal{Z}(p)$ times its corresponding part in *cost*$(\boldsymbol{x})$.

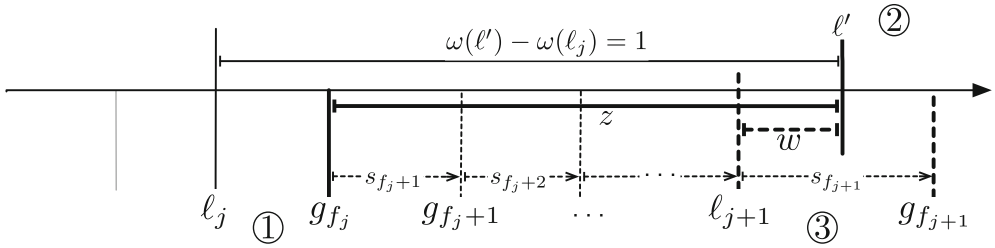

Fig. 2. The proof of $\mathbf{E}\big[\omega(\ell_{j+1}) - \omega(\ell_j) \mid S_j\big] \geq \mathcal{Z}(p)$. Dashed lines are for quantities that are independent of the current state S_j, but determined by the next state S_{j+1}.

The global order cost $C|\boldsymbol{g}|$. The expectation of each global cutoff g_{i+1}, given g_i, is $g_i + \mathbf{E}[p]$, which (by definition of $\mathcal{Z}(p)$) is at least $g_i + \mathcal{Z}(p)$. The final index I is the first such that $g_I \geq \hat{U} - 1$, so $g_I < \hat{U}$. By Wald's equation (Lemma 5), since I is a stopping time, the expected length of $\boldsymbol{g}$ is at most $\hat{U}/\mathcal{Z}(p)$. So, $\mathbf{E}[C|\boldsymbol{g}|]$ is at most $C\hat{U}/\mathcal{Z}(p)$. In comparison the global order cost in *cost*($\boldsymbol{x}$) is $C\sum_{t=1}^{U} x_t = C\hat{U}$.

The retailer cost $c_\rho|\boldsymbol{\ell}^\rho|$ *for* ρ. Fix a retailer ρ. Since ρ is fixed for the rest of the proof, we may omit it as a subscript or superscript. Let $\boldsymbol{\ell}$ be ρ's local cutoff sequence $(\ell_1, \ell_2, \ldots, \ell_J)$. For each $j = 1, 2, \ldots, J$, define the *state* S_j *after step* j to be the first f_j random samples, where f_j is the number of random samples needed to determine ℓ_j. Generally, a given global cutoff g_i will be chosen as the jth local cutoff ℓ_j iff $\omega(g_i) - \omega(\ell_{j-1}) \leq 1 < \omega(g_{i+1}) - \omega(\ell_{j-1})$. So, f_j equals $i + 1$, where i is the index such that $g_i = \ell_j$. That is, g_{f_j} *follows* ℓ_j in the global sequence. (The only exception is the *last* local cutoff ℓ_J, which can be the maximum global cutoff g_I, in which case it is not followed by any cutoff and $f_J = I$.)

For the analysis, define $S_0 = (s_1)$ and $\ell_0 = 0$ so $\omega(\ell_0) = 0$.

We claim that, *with each step* $j = 0, \ldots, J-1$, *given the state* S_j *after step* j, *the expected increase in* $\omega(\cdot)$ *during step* $j+1$, *namely* $\mathbf{E}[\omega(\ell_{j+1}) - \omega(\ell_j) \mid S_j]$, *is at least* $\mathcal{Z}(p)$. Before we prove the claim, note that this implies the desired bound: by the stopping condition for $\boldsymbol{\ell}$, the total increase in ω is at most $\omega(\hat{U})$, so by Wald's equation (using that the last index J is a stopping time), the expectation of J is at most $\omega(\hat{U})/\mathcal{Z}(p)$. So, $\mathbf{E}[\,c_\rho|\boldsymbol{\ell}|\,]$, the expected cost for retailer ρ, is at most $c_\rho\,\omega(\hat{U})/\mathcal{Z}(p)$. In comparison, the cost for retailer ρ in *cost*($\boldsymbol{x}$) is $c_\rho \sum_{t=1}^{U} x_t^\rho = c_\rho\,\omega(\hat{U})$.

To prove the claim, we describe how moving from state S_j to state S_{j+1} is a play of the tally game in the definition of $\mathcal{Z}(p)$. Fix any j and condition on the state S_j. Fig. 2 shows the steps:

① The current state S_j determines the j'th local cutoff ℓ_j and the following global cutoff g_{f_j}.

② Given ℓ_j and g_{f_j}, the next local cutoff for retailer ρ will be the maximum global cutoff in the interval $[g_{f_j}, \ell']$, where ℓ' is chosen so that $\omega(\ell') - \omega(\ell_j)$ equals 1. (Note that $\ell' < \hat{U}$ because, since we haven't stopped yet, $\omega(\hat{U}) - \omega(\ell_j) > 1$.)

③ The algorithm reaches the next state S_{j+1} by drawing some more random samples $s_{f_j+1}, s_{f_j+2}, \dots, s_i$ from p, stopping with the first index i such that the corresponding global cutoff g_i exceeds ℓ'. (That is, such that $g_i = g_{f_j} + s_{f_j+1} + \dots + s_i > \ell'$.) The next local cutoff ℓ_{j+1} is g_{i-1} (the global cutoff just before g_i, so that $g_{i-1} \le \ell' < g_i$) and this index i is f_{j+1}; that is, the next state S_{j+1} is $(s_1, s_2, \dots, s_i)$.

Step ③ is a play of the "tally game" in the definition of $\mathcal{Z}(p)$, with threshold $z = \ell' - g_{f_j}$. The waste w equals the gap $\ell' - \ell_{j+1}$. By the definition of $\mathcal{Z}(p)$, the expectation of w is $\mathcal{W}(p, z) \le 1 - \mathcal{Z}(p)$. Finally,

$$\omega(\ell_{j+1}) - \omega(\ell_j) \;=\; 1 - (\omega(\ell') - \omega(\ell_{j+1})) \;\ge\; 1 - (\ell' - \ell_{j+1}) \;=\; 1 - w.$$

The expectation of $1 - w$ is at least $\mathcal{Z}(p)$, proving the claim.

The careful reader may notice that the above analysis is incorrect for the last step J, because it may happen that there is no global cutoff after ℓ'. (Then $\ell_J = g_I = \max_i g_i$.) To analyze this case, imagine modifying the algorithm so that, in choosing $\boldsymbol{g} = (g_1, g_2, \dots, g_I)$, instead of stopping with $I = i$ such that $g_i \ge \hat{U} - 1$, it stops with $I = i$ such that $g_i \ge \hat{U}$. Because the last global cutoff is now at least $\hat{U}$, and $\ell' < \hat{U}$, there is always a global cutoff after ℓ'. So the previous analysis is correct for the modified algorithm, and its expected local order cost is bounded as claimed. To finish, observe that, since this modification only *extends* $\boldsymbol{g}$, it cannot *decrease* the number of local cutoffs selected from $\boldsymbol{g}$, so the modification does not decrease the local order cost. □

2.1 Upper Bound of $e/(e-1) \approx 1.582$

The next utility lemma says that, in analyzing the expected waste in the tally game, it is enough to consider thresholds z in $[0, 1]$.

Lemma 2. *For any distribution p on $[0,1]$, $\sup_{z \ge 0} \mathcal{W}(p, z) \;=\; \sup_{z \in [0,1)} \mathcal{W}(p, z)$.*

Proof. Play the tally game with any threshold $z > 1$. Consider the first prefix sum $s_1 + s_2 + \dots + s_h$ of the samples such that the "slack" $z - (s_1 + s_2 + \dots + s_h)$ is at most 1. Let random variable z' be this slack. Note $z' \in [0, 1)$. Then, conditioned on $z' = y$, the expected waste is $\mathcal{W}(p, y)$, which is at most $\sup_{Y \in [0,1]} \mathcal{W}(p, Y)$. Thus, for *any* threshold $z \ge 1$, $\mathcal{W}(p, z)$ is at most $\sup_{Y \in [0,1)} \mathcal{W}(p, Y)$. □

Now consider the specific probability distribution p on $[0, 1]$ with probability density function $p(y) = 1/y$ for $y \in [1/e, 1]$ and $p(y) = 0$ elsewhere.

Lemma 3. *For this distribution p, $\mathcal{Z}(p) \;\ge\; (e-1)/e \;=\; 1 - 1/e$.*

Proof. By Lemma 2, $\mathcal{Z}(p)$ is the minimum of $\mathbf{E}[p]$ and $1 - \max_{z \in [0,1]} \mathcal{W}(p, z)$.

By direct calculation, $\mathbf{E}[p] = \int_{1/e}^{1} y\, p(y)\, dy \;=\; \int_{1/e}^{1} 1\, dy = 1 - 1/e$.

Now consider playing the tally game with threshold z. If $z \in [0, 1/e]$, then (since the waste is at most z) trivially $\mathcal{W}(p, z) \le z \le 1/e$.

So consider any $z \in [1/e, 1]$. Let s_1 be just the first sample.

The waste is z if $s_1 > z$ and otherwise is at most $z - s_1$.
So, the expected waste is at most $\Pr[s_1 > z]z + \Pr[s_1 \le z]\,\mathbf{E}[z - s_1 \mid s_1 \le z]$.
This simplifies to $z - \Pr[s_1 \le z]\,\mathbf{E}[s_1 \mid s_1 \le z]$, which by calculation is

$$z - \int_{1/e}^{z} y\,p(y)\,dy \;=\; z - \int_{1/e}^{z} dy \;=\; z - (z - 1/e) \;=\; 1/e. \qquad \square$$

2.2 Upper Bound of 1.574

Next we define a probability distribution on $[0,1]$ that has a point mass at 1. Fix $\theta = 0.36455$ (slightly less than $1/e$). Over the half-open interval $[0,1)$, the probability density function p is $p(y) = \begin{cases} 0 & \text{for } y \in [0,\theta) \\ 1/y & \text{for } y \in [\theta, 2\theta) \\ \frac{1-\ln((y-\theta)/\theta)}{y} & \text{for } y \in [2\theta, 1). \end{cases}$
The probability of choosing 1 is $1 - \int_0^1 p(y)\,dy \approx 0.0821824$.

Note that $p(y) \ge 0$ for $y \in [2\theta, 1)$ since $\ln((1-\theta)/\theta) \approx 0.55567$.

Lemma 4. *The statistic $\mathcal{Z}(p)$ for this p is at least $0.63533 > 1/1.574$.*

The proof is in the full paper [4]. Here is a sketch.

Proof (sketch). By Lemma 2, $\mathcal{Z}(p)$ is the minimum of $\mathbf{E}[p]$ and $1 - \sup_{z\in[0,1]} \mathcal{W}(p,z)$.

That $\mathbf{E}[p] \ge 0.63533$ follows from a direct calculation (about five lines; details in the full proof).

It remains to show $1 - \sup_{z\in[0,1)} \mathcal{W}(p,z) \ge 0.63533$. To do this, we show $\sup_{z\in[0,1)} \mathcal{W}(p,z) = \theta$ ($\le 1 - 0.63533$).

In the tally game defining $\mathcal{W}(p,z)$, let s_1 be the first random sample drawn from p. If $s_1 > z$, then the waste equals z. Otherwise, the process continues recursively with z replaced by $z' = z - s_1$. This gives the recurrence

$$\mathcal{W}(p,z) \;=\; z\,\Pr[s_1 > z] + \int_\theta^z \mathcal{W}(p, z-y)\,p(y)\,dy.$$

We analyze the right-hand side of the recurrence in three cases.

Case (i) $z \in [0,\theta)$. The recurrence gives $\mathcal{W}(p,z) = z$ because $\Pr[s_1 > z] = 1$.

Case (ii) $z \in [\theta, 2\theta)$. For $y \in [\theta, z]$, we have $z - y < \theta$, so by Case (i) $\mathcal{W}(p, z-y) = z - y$. Substituting and calculating (about two lines) gives $\mathcal{W}(p,z) = \theta$.

Case (iii) $z \in [2\theta, 1)$. For $y \in [\theta, z]$, we have $z - y < 2\theta$, so Case (i) or (ii) applies to simplify $\mathcal{W}(p, z-y)$ (to $z-y$ if $z - y < \theta$ or θ otherwise). The calculation (about seven lines) gives $\mathcal{W}(p,z) = \theta$. $\square$

Theorem 1. *JRP-D has a randomized polynomial-time 1.574-approximation algorithm, and the integrality gap of the LP relaxation is at most 1.574.*

Proof. By Lemma 4, for any fractional solution $\boldsymbol{x}$, the algorithm Round_p (using the probability distribution p from that lemma) returns a feasible schedule of expected cost at most 1.574 times $cost(\boldsymbol{x})$.

To see that the schedule can be computed in polynomial time, note first that the (discrete-time) LP relaxation can be solved in polynomial time. The optimal solution $\boldsymbol{x}$ is minimal, so each x_t is at most 1, so $\hat{U} = \sum_t x_t$ is at most the number of demands, n. In the algorithm Round_p, each sample from the distribution p from Lemma 4 can be drawn in polynomial time. Each sample is $\Omega(1)$, and the sum of the samples is at most $\hat{U} \leq n$, so the number of samples is $O(n)$. Then, for each retailer ρ, each integral $\mu(\ell^\rho_{j-1}, g)$ in step 3 can be evaluated in constant amortized time per evaluation, so the time per retailer is $O(n)$. □

For the record, here is the variant of Wald's equation (also known as Wald's identity or Wald's lemma, and a consequence of standard "optional stopping" theorems) that we use above. Consider a random experiment that, starting from a fixed start state S_0, produces a random sequence of states $S_1, S_2, S_3, \ldots$ Let random index $T \in \{0, 1, 2, \ldots\}$ be a stopping time for the sequence, that is, for any positive integer t, the event "$T < t$" is determined by state S_t. Let function $\phi : \{S_t\} \to \mathbb{R}$ map the states to $\mathbb{R}$.

Lemma 5 (Wald's equation). *Suppose that*
(i) $(\forall t < T)\ \mathbf{E}[\phi(S_{t+1}) \mid S_t] \geq \phi(S_t) + \xi$ *for fixed* ξ*, and*
(ii) either $(\forall t < T)\ \phi(S_{t+1}) - \phi(S_t) \geq F$ *or* $(\forall t < T)\ \phi(S_{t+1}) - \phi(S_t) \leq F$*, for some fixed finite* F*, and* T *has finite expectation.*

Then $\xi\, \mathbf{E}[T] \leq \mathbf{E}[\phi(S_T) - \phi(S_0)]$.

The proof is standard; it is in the full paper [4].

In the applications here, we always have $\xi = \mathcal{Z}(p) > 0$ and $\phi(S_T) - \phi(S_0) \leq U$ for some fixed U. In this case Wald's equation implies $\mathbf{E}[T] \leq U/\mathcal{Z}(p)$.

3 Upper Bound of 1.5 for Equal-Length Periods

In this section, we present a 1.5-approximation algorithm for the case where all the demand periods are of equal length. For convenience, we allow here release times and deadlines to be rational numbers and we assume that all demand periods have length 1.

We denote the input instance by $\mathcal{I}$. Let the *width* of an instance be the difference between the deadline of the last demand and the release time of the first one. The building block of our approach is an algorithm that creates an optimal solution to an instance of width at most 3. Later, we divide $\mathcal{I}$ into overlapping sub-instances of width 3, solve each of them optimally, and finally show that by aggregating their solutions we obtain a 1.5-approximation for $\mathcal{I}$.

Lemma 6. *A solution to any instance* $\mathcal{J}$ *of width at most* 3 *consisting of unit-length demand periods can be computed in polynomial time.*

Proof. We shift all demands in time, so that $\mathcal{J}$ is entirely contained in interval $[0, 3]$. Recall that C is the warehouse ordering cost and c_ρ is the ordering cost of retailer $\rho \in \{1, 2, ..., m\}$. Without loss of generality, we can assume that all retailers $1, .., m$ have some demands.

Let $d_{\min}$ be the first deadline of a demand from $\mathcal{J}$ and $r_{\max}$ the last release time. If $r_{\max} \leq d_{\min}$, then placing one order at any time from $[r_{\max}, d_{\min}]$ is sufficient (and necessary). Its cost is then equal to $C + \sum_\rho c_\rho$.

Thus, in the following we focus on the case $d_{\min} < r_{\max}$. Any feasible solution has to place an order at or before $d_{\min}$ and at or after $r_{\max}$. Furthermore, by shifting these orders we may assume that the first and last orders occur exactly at times $d_{\min}$ and $r_{\max}$, respectively.

The problem is thus to choose a set T of warehouse ordering times that contains $d_{\min}$, $r_{\max}$, and possibly other times from the interval $(d_{\min}, r_{\max})$, and then to decide, for each retailer ρ, which warehouse orders it joins. Note that $r_{\max} - d_{\min} \leq 1$, and therefore each demand period contains $d_{\min}$, $r_{\max}$, or both. Hence, all demands of a retailer ρ can be satisfied by joining the warehouse orders at times $d_{\min}$ and $r_{\max}$ at additional cost of $2b_\rho$. It is possible to reduce the retailer ordering cost to c_ρ if (and only if) there is a warehouse order that occurs within $\mathcal{D}_\rho$, where $\mathcal{D}_\rho$ is the intersection of all demand periods of retailer ρ. (To this end, $\mathcal{D}_\rho$ has to be non-empty.)

Hence, the optimal cost for $\mathcal{J}$ can be expressed as the sum of four parts:

(i) the unavoidable ordering cost c_ρ for each retailer ρ,
(ii) the additional ordering cost c_ρ for each retailer ρ with empty $\mathcal{D}_\rho$,
(iii) the total warehouse ordering cost $C \cdot |T|$, and
(iv) the additional ordering cost c_ρ for each retailer ρ whose $\mathcal{D}_\rho$ is non-empty and does not contain any ordering time from T.

As the first two parts of the cost are independent of T, we focus on minimizing the sum of parts (iii) and (iv), which we call the *adjusted cost*. Let $AC(t)$ be the minimum possible adjusted cost associated with the interval $[d_{\min}, t]$ under the assumption that there is an order at time t. Formally, $AC(t)$ is the minimum, over all choices of sets $T \subseteq [d_{\min}, t]$ that contain $d_{\min}$ and t, of $C \cdot |T| + \sum_{\rho \in Q(T)} c_\rho$, where $Q(T)$ is the set of retailers ρ for which $\mathcal{D}_\rho \neq \emptyset$ and $\mathcal{D}_\rho \subseteq [0, t] - T$. (Note that the second term consists of expenditures that actually occur outside the interval $[d_{\min}, t]$.)

As there are no $\mathcal{D}_\rho$'s strictly to the left of $d_{\min}$, $AC(d_{\min}) = C$. Furthermore, to compute $AC(t)$ for any $t \in (d_{\min}, r_{\max}]$, we can express it recursively using the value of $AC(u)$ for $u \in [d_{\min}, t)$ being the warehouse order time immediately preceding t in the set T that realizes $AC(t)$. This gives us the formula

$$AC(t) = C + \min_{u \in [d_{\min}, t)} \left(AC(u) + \sum_{\rho : \emptyset \neq \mathcal{D}_\rho \subset (u,t)} c_\rho \right) .$$

In the minimum above, we may restrict computation of $AC(t)$ to t's and u's that are ends of demand periods. Hence, the actual values of function $AC(\cdot)$ can be computed by dynamic programming in polynomial time. Finally, the total adjusted cost is equal to $AC(r_{\max})$. Once we computed the minimum adjusted cost, recovering the actual orders can be performed by a straightforward extension of the dynamic programming presented above. □

Now, we construct an approximate solution for the original instance $\mathcal{I}$ consisting of unit-length demand periods. For $i \in \mathbb{N}$, let $\mathcal{I}_i$ be the sub-instance containing all demands entirely contained in $[i, i+3)$. By Lemma 6, an optimal solution for $\mathcal{I}_i$, denoted $A(\mathcal{I}_i)$, can be computed in polynomial time. Let S_0 be the solution created by aggregating $A(\mathcal{I}_0), A(\mathcal{I}_2), A(\mathcal{I}_4), \ldots$ and S_1 by aggregating $A(\mathcal{I}_1), A(\mathcal{I}_3), A(\mathcal{I}_5), \ldots$. Among solutions S_0 and S_1, we output the one with the smaller cost.

Theorem 2. *The above algorithm produces a feasible solution of cost at most* 1.5 *times the optimum cost.*

Proof. Each unit-length demand of instance $\mathcal{I}$ is entirely contained in some $\mathcal{I}_{2k}$ for some $k \in \mathbb{N}$. Hence, it is satisfied in $A(\mathcal{I}_{2k})$, and thus also in S_0, which yields the feasibility of S_0. An analogous argument shows the feasibility of S_1.

To estimate the approximation ratio, we fix an optimal solution Opt for instance $\mathcal{I}$ and let opt_i be the cost of Opt's orders in the interval $[i, i+1)$. Note that Opt's orders in $[i, i+3)$ satisfy all demands contained entirely in $[i, i+3)$. Since $A(\mathcal{I}_i)$ is an optimal solution for these demands, we have $cost(A(\mathcal{I}_i)) \le opt_i + opt_{i+1} + opt_{i+2}$ and, by taking the sum, we obtain $cost(S_0) + cost(S_1) \le \sum_i cost(A(\mathcal{I}_i)) \le 3 \cdot cost(\text{Opt})$. Therefore, either of the two solutions (S_0 or S_1) has cost at most $1.5 \cdot cost(\text{Opt})$. □

4 Two Lower Bounds

In this section we present two lower bounds on the integrality gap of the LP relaxation from Section 1:

Theorem 3. (i) *The integrality gap of the LP relaxation is at least* $\frac{1}{2}(1+\sqrt{2})$, *which is at least* 1.207. (ii) *The integrality gap is at least* 1.2 *for instances with equal-length demand periods.*

In the full paper [4], we sketch how the lower bound in Part (i) can be increased to 1.245 via a computer-based proof; we also give the complete proof of Thm. 3 (about six pages altogether). Here is a sketch of that proof.

Proof (sketch). It is convenient to work with a *continuous-time* variant of the LP, in which the universe $\mathcal{U}$ of allowable release times, deadlines and order times is the entire interval $[0, U]$, where U is the maximum deadline. Time t now is a real number ranging over interval $[0, U]$. A fractional solution is now represented by *functions*, $x : [0, U] \to \mathbb{R}_{\ge 0}$ and $x^\rho : [0, U] \to \mathbb{R}_{\ge 0}$, for each retailer ρ. To retain consistency, we will write x_t and x_t^ρ for the values of these functions. For any fractional solution $\boldsymbol{x}$, then, in the LP formulation each sum over t is replaced by the appropriate integral. For example, the objective function will now take form $\int_{t=0}^{U} (C\, x_t + \sum_{\rho=1}^{m} c_\rho\, x_t^\rho)$. By a straightforward limit argument (to be provided in the final version of the paper), the continuous-time LP has the same integrality gap as the discrete-time LP.

(i) The instance used to prove Part (i) has $C = 1$ and two retailers numbered (for convenience) 0 and 1, one with $c_0 = 0$ and the other with $c_1 = \sqrt{2}$. We use infinitely many demand periods: for any t, the first retailer has demand periods $[t, t+1]$ and the second retailer has demand periods $[t, t+\sqrt{2}]$. A fractional solution where retailer 0 orders at rate 1 and retailer 1 orders at rate $1/\sqrt{2}$

is feasible and its cost is 2 per time step. Now consider some integral solution. Without loss of generality, retailer 0 orders any time a warehouse order is issued. Retailer 0 must make at least one order per time unit, so his cost (counting the warehouse order cost as his) is 1 per time unit. Retailer 1 must make at least one order in any time interval of length $\sqrt{2}$, so the cost of his orders, not including the warehouse cost, is at least 1 per time unit as well. This already gives us cost 2 per time unit, the same as the optimal fractional cost. But in order to synchronize the orders of retailer 1 with the warehouse orders, the schedule needs to increase either the number of retailer 1's orders or the number of warehouse orders by a constant fraction, thus creating a gap.

(ii) The argument for Part (ii) is more involved. We only outline the general idea. Take $C = 1$ and three retailers numbered 0, 1 and 2, each with order cost $c_\rho = \frac{1}{3}$. The idea is to create an analogue of a 3-cycle, which has a fractional vertex cover with all vertices assigned value $\frac{1}{2}$ and total cost $\frac{5}{6}$, while any integral cover requires two vertices. We implement this idea by starting with the following fractional solution x: if $t \bmod 3 = 0$, then $x_t = x_t^0 = x_t^1 = \frac{1}{2}$ and $x_t^2 = 0$; if $t \bmod 3 = 1$, then $x_t = x_t^1 = x_t^2 = \frac{1}{2}$ and $x_t^0 = 0$; if $t \bmod 3 = 2$, then $x_t = x_t^0 = x_t^2 = \frac{1}{2}$ and $x_t^1 = 0$. The cost is $\frac{5}{6}$ per time unit. Then we choose demand periods that x satisfies, but such that, in any (integral) schedule, each retailer must have at least one order in every three time units $\{t, t+1, t+3\}$, *and* there has to be a warehouse order in every two time units $\{t, t+1\}$. These costs independently add up to $\frac{5}{6}$ per time unit, even ignoring the requirement that retailers have orders only when the warehouse does. To synchronize the orders to meet this additional requirement, any schedule must further increase the order rate by a constant fraction, thus creating a gap. □

5 APX Hardness for Unit Demand Periods

Theorem 4. *JRP-D is $\mathbb{APX}$-hard even if restricted to instances with unit demand periods and with at most four demands per retailer.*

The proof (about four pages) is in the full paper [4]. Here is the idea.

Proof (idea). We use the result by Alimonti and Kann [1] that Vertex Cover is $\mathbb{APX}$-hard even for cubic graphs. For any given cubic graph $G = (V, E)$ with n vertices (where n is even) and $m = 1.5n$ edges, in polynomial time we construct an instance $\mathcal{J}_G$ of JRP-D, such that the existence of a vertex cover for G of size at most K is equivalent to the existence of an order schedule for $\mathcal{J}_G$ of cost at most $10.5n + K + 6$. In $\mathcal{J}_G$ all demand periods have the same length and each retailer has at most four demands. The construction consists of gadgets that represent G's vertices and edges. The main challenge, related to the equal-length restriction, is in "transmitting information" along the time axis about the vertices chosen for a vertex cover. We resolve it by having each vertex represented twice and assuring consistency via an appropriate sub-gadget. □

6 Final Comments

The integrality gap for standard JRP-D LP relaxation is between 1.245 and 1.574. We conjecture that neither bound is tight, but that the refined distribution for the tally game given here is essentially optimal, so that improving the upper bound will require a different approach.

There is a simple algorithm for JRP-D that provides a $(1, 2)$-approximation, in the following sense: its warehouse order cost is not larger than that in the optimum, while its retailer order cost is at most twice that in the optimum [12]. One can then try to balance the two approaches by choosing each algorithm with a certain probability. This simple approach does not improve the ratio. But it may be possible to achieve a better ratio if, instead of using our algorithm as presented, we appropriately adjust the probability distribution.

If we parametrize JRP-D by the maximum number p of demand periods of each retailer, its complexity status is essentially resolved: for $p \geq 3$ the problem is $\mathbb{APX}$-hard [12], while for $p \leq 2$ it can be solved in polynomial time (by a greedy algorithm for $p = 1$ and dynamic programming for $p = 2$). In case of equal-length demand periods, we showed that the problem is $\mathbb{APX}$-hard for $p \geq 4$, but the case $p = 3$ remains open, and it would be nice to settle this case as well. We conjecture that in this case the problem is $\mathbb{NP}$-complete.

Acknowledgements. We would like to thank Łukasz Jeż, Jiří Sgall, and Grzegorz Stachowiak for stimulating discussions and useful comments.

References

1. Alimonti, P., Kann, V.: Some APX-completeness results for cubic graphs. Theoretical Computer Science 237(1-2), 123–134 (2000)
2. Arkin, E., Joneja, D., Roundy, R.: Computational complexity of uncapacitated multi-echelon production planning problems. Operations Research Letters 8(2), 61–66 (1989)
3. Becchetti, L., Marchetti-Spaccamela, A., Vitaletti, A., Korteweg, P., Skutella, M., Stougie, L.: Latency-constrained aggregation in sensor networks. ACM Transactions on Algorithms 6(1), 13:1–13:20 (2009)
4. Bienkowski, M., Byrka, J., Chrobak, M., Dobbs, N., Nowicki, T., Sviridenko, M., Świrszcz, G., Young, N.E.: Approximation algorithms for the joint replenishment problem with deadlines. CoRR abs/1212.3233v2 (2013), http://arxiv.org/abs/1212.3233v2
5. Brito, C., Koutsoupias, E., Vaya, S.: Competitive analysis of organization networks or multicast acknowledgement: How much to wait? In: Proc. of the 15th ACM-SIAM Symp. on Discrete Algorithms (SODA), pp. 627–635 (2004)
6. Buchbinder, N., Kimbrel, T., Levi, R., Makarychev, K., Sviridenko, M.: Online make-to-order joint replenishment model: Primal dual competitive algorithms. In: Proc. of the 19th ACM-SIAM Symp. on Discrete Algorithms (SODA), pp. 952–961 (2008)
7. Khanna, S., Naor, J.(S.), Raz, D.: Control message aggregation in group communication protocols. In: Widmayer, P., Triguero, F., Morales, R., Hennessy, M., Eidenbenz, S., Conejo, R. (eds.) ICALP 2002. LNCS, vol. 2380, pp. 135–146. Springer, Heidelberg (2002)

8. Levi, R., Roundy, R., Shmoys, D.B.: A constant approximation algorithm for the one-warehouse multi-retailer problem. In: Proc. of the 16th ACM-SIAM Symp. on Discrete Algorithms (SODA), pp. 365–374 (2005)
9. Levi, R., Roundy, R., Shmoys, D.B.: Primal-dual algorithms for deterministic inventory problems. Mathematics of Operations Research 31(2), 267–284 (2006)
10. Levi, R., Roundy, R., Shmoys, D.B., Sviridenko, M.: A constant approximation algorithm for the one-warehouse multiretailer problem. Management Science 54(4), 763–776 (2008)
11. Levi, R., Sviridenko, M.: Improved approximation algorithm for the one-warehouse multi-retailer problem. In: Díaz, J., Jansen, K., Rolim, J.D.P., Zwick, U. (eds.) APPROX 2006 and RANDOM 2006. LNCS, vol. 4110, pp. 188–199. Springer, Heidelberg (2006)
12. Nonner, T., Souza, A.: Approximating the joint replenishment problem with deadlines. Discrete Mathematics, Algorithms and Applications 1(2), 153–174 (2009)